DOCUMENT RESUME

ED 204 364 TM 810 345

AUTHOR	Walter, L. James; And Others
TITLE	Testing Programs in Schools: What's New? A Special Issue.
INSTITUTION	Nebraska Association for Supervision and Curriculum Development.; Nebraska Council of School Administrators, Lincoln.
PUB DATE	81
NOTE	60p.
JOURNAL CIT	Catalyst, The Journal of the Nebraska Council of School Administrators; v4 n2 Win 1981
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Criterion Referenced Tests; *Information Utilization; Latent Trait Theory; Program Improvement; *School Districts; Scores; *Standardized Tests; State Surveys; *Testing; *Testing Programs; Test Selection; Test Validity
IDENTIFIERS	Fremont Public School NE; Mental Measurements Yearbook; *Nebraska; Omaha Public Schools NE; Rasch Model; *Test Use

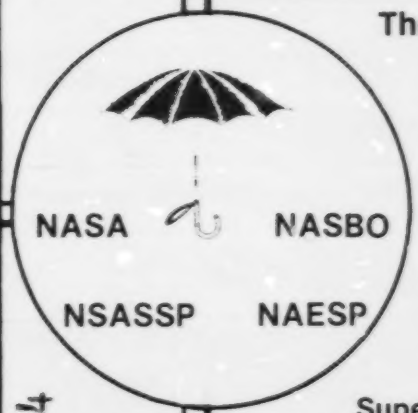
ABSTRACT

This special issue of "Catalyst", entitled Testing Programs in Schools: What's New?, was published jointly by the Nebraska Council of School Administrators and the Nebraska Association for Supervision and Curriculum. The articles emphasize desirability of designing testing programs to meet the information needs of educators and the community. Articles included are: a guest editorial, Toward Improved Testing Programs, by L. James Walter; Testing Programs in Nebraska School Districts, by Conrad Carlberg; So, You're Going to Buy Another Standardized Test... Whose? Why?? Who Cares???, by H. James Funk and Charles R. DuVall; Practical Uses and Interpretations of Achievement Testing, by Paul G. Monkowski; A Helping Hand in Testing: The Mental Measurements Yearbook, by James V. Mitchell, Jr.; Evaluating Student Achievement: New Technology Applied to Old Measurement Problems, by Mardel R. Kolls; So You Want to Get a Student Tested? Tips for the Building Administrator, by James M. Boland; A Review of a Testing Program: The Fremont Public Schools, by James Windeshausen; and The Testing Program of the Omaha Public Schools, by Paul J. Malcom. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

CATALYST

The Journal of the Nebraska Council of School Administrators



A SPECIAL ISSUE

Published In Cooperation With

The Nebraska Association for

Supervision and Curriculum Development

TESTING PROGRAMS IN SCHOOLS: WHAT'S NEW?

ED204364

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 810 315

Volume 4 Winter, 1981 Number 2

Catalyst
The Journal of the Nebraska Council of School Administrators

NCSA Executive Committee

John Lammel, Chairman, (Past President, NSASSP), Principal, Millard Senior High School, 14905 Q Street, Omaha, 68137; **Harold Koch,** Vice Chairman, (President, NASA), Superintendent, Gering City Schools, 800 Q Street, Gering, 69341; **Nyla Alexander,** Secretary, (President, NAESP), Principal, Wasmer Elementary School, 1613 West Division, Grand Island, 68801. **Members: Fred Bellum,** Superintendent, (President-Elect, NASA), Columbus City Schools, Box 947, Columbus 68601; **Harold Bennett,** Superintendent, (Past President, NASA), McCook City Schools, 700 West 7th Street, McCook, 69001; **John Deegan,** Principal, (Past President, NAESP), Peter Sarpy Elementary, 30th and Vandenberg, Omaha, 68123; **Bob Duryea,** Business Manager, (President-Elect, NASBO), Alliance City Schools, 1450 Box Butte, Alliance, 69301; **Les George,** Principal, (President-Elect, NSASSP), Benson High School, 5120 Maple Street, Omaha, 68104; **Ruth Leopold,** Business Manager, (Past President, NASBO), McCook City Schools, 700 West 7th Street, McCook, 69001; **Gerald Menke,** Principal, (President, NSASSP), Kearney Junior High, 915 West 35th Street, Kearney, 68847; **Stanley Wilcox,** Asst. Superintendent, (President, NASBO), Papillion LaVista Public Schools, 420 South Washington, Papillion, 68046; **Doug Townsend,** Principal, (President-Elect, NAESP), Lincoln Elementary, 2101 So. 15th Street, Plattsmouth, 68048. **Ex-Officio Members: Anne Campbell,** Commissioner, State Department of Education, 301 Centennial Mall South, Lincoln, 68509; **O.W. Kopp,** Director, Center for Curriculum and Instruction, 202 Teachers College, University of Nebraska-Lincoln, Lincoln, 68588; **Robert Egbert,** Dean, Teachers College, 101 Teachers College, University of Nebraska-Lincoln, Lincoln, 68588; **Justin King,** Executive Director, Nebraska State School Board Association, 521 South 14th Street, Lincoln, 68588; **Martha Fricke,** 408 North 12th Street, Ashland, 68003.

NCSA Communications Committee

Norbert Schuerman, Chairman, (Omaha Public Schools, 3902 Davenport Street, Omaha, 68131); **Paul Baker,** (Fremont Junior High School, 130 East 9th Street, Fremont, 68025); **Curt Crandall,** (Sheridan Elementary School, 3100 Plymouth Street, Lincoln, 68502); **Steve Deger,** (Humphrey Public Schools, Box 278, Humphrey, 68642); **Larry Dlugosh,** (Elkhorn Public Schools, 502 Glenn, Elkhorn, 68022); **Sally Ganem,** (Gretna Elementary School, 802 South Street, Gretna, 68028); **Barry Goddard,** (Weeping Water Public Schools, Weeping Water, 68463); **Nancy Holloran,** (Millard Public Schools, 12805 L Street, Omaha, 68137); **Edwin Johnson,** (Syracuse Dunbar Avoca Schools, 5550 7th Street, Syracuse, 68446); **Edgar A. Kelley,** (University of Nebraska-Lincoln, 105-D Henzlik Hall, Lincoln, 68588); **Bill Kenagy,** (Holdrege Junior High, 1006th Street, Holdrege, 69361); **Harley Rector,** (Norfolk Junior High, 5th and Pasewalk, Norfolk, 68701); **Paul Toms,** (School District of York, 611 Platte Avenue, York, 68467). **Ex-Officio: Robert Peterson,** NCSA Executive Secretary; **John Lammel,** Chairman, NCSA Executive Committee.

NCSA Executive Secretary

Robert Peterson, Executive Secretary, Nebraska Council of School Administrators, First National Bank Building, Suite 1704, 233 South 13th Street, Lincoln, 68508 (Telephone: 402-476-8055).

Editorial Staff

Edgar A. Kelley, Editor, 105-D Henzlik Hall, University of Nebraska-Lincoln, Lincoln, 68588 (Telephone: 402-472-1985); **Frederick C. Wendel,** Co-Editor, 36 Henzlik Hall, University of Nebraska-Lincoln, Lincoln 68588 (Telephone: 402-472-3726); **Robert W. Shannon,** Assistant Editor, 104-B Henzlik Hall, University of Nebraska-Lincoln, Lincoln, 68588 (Telephone: 402-472-1986).

Publication Information

Catalyst is published in the Fall, Winter, Spring and Summer by the Nebraska Council of School Administrators in cooperation with Teachers College of the University of Nebraska-Lincoln. Subscription rate, $15 per year. Single copies, $4 each. Remit payment for subscriptions or single copies with order. All orders should be sent to the NCSA Executive Secretary. All advertising copy and manuscripts should be sent to the Editor. All purchase orders and payment for advertising should be sent to the NCSA Executive Secretary. Member, Educational Press Association of America. Entered as second-class matter at the post office, Lincoln, Nebraska. Opinions expressed in Catalyst do not necessarily represent positions adopted by the Nebraska Council of Administrators except in those instances where a position is specifically identified as having been adopted by the NCSA.

Copyright 1981 by the Nebraska Council of School Administrators

CATALYST

The Journal of the Nebraska Council of School Administrators
Volume 4, Number 2 Winter, 1981

TESTING PROGRAMS IN SCHOOLS: WHAT'S NEW?

A Special Issue Published in Cooperation With the Nebraska Association for Supervision and Curriculum Development.

Contents

The Editor's Desk — Edgar A. Kelley

Guest Editorial: Toward Improved Testing Programs — L. James Walter

Testing Programs In Nebraska School Districts — Conrad Carlberg

So, You're Going To Buy Another Standardized Test... Whose? Why?? Who Cares??? — H. James Funk, Charles R. DuVall

Practical Uses And Interpretations Of Achievement Testing — Paul G. Monkowski

A Helping Hand In Testing: The Mental Measurements Yearbook — James V. Mitchell, Jr.

Evaluating Student Achievement: New Technology Applied to Old Measurement Problems — Mardel R. Kolls

So You Want To Get A Student Tested? Tips For The Building Administrator — James M. Boland

A Review Of A Testing Program: The Fremont Public Schools — James Windeshausen

The Testing Program Of The Omaha Public Schools — Paul J. Malcom

THE EDITOR'S DESK

About This Issue. This issue represents a "first" for **Catalyst;** it is the first issue of the journal prepared in cooperation with another professional organization. The co-publisher of this issue is the Nebraska Association for Supervision and Curriculum Development and NASCD Executive Secretary, L. James Walter, served as the guest editor for this issue. The cooperation of NASCD and, in particular, the cooperation of Jim Walter is appreciated. To all NASCD members and to Jim, thank you on behalf of NCSA and its members.

A Point of View. Reading the articles in this issue of **Catalyst** reminds one in subtle, and not-so-subtle, ways that educators can become too removed from the citizens we serve. In many of our practices, including testing programs but certainly not limited to that example, we place our concern almost solely on information which will serve our immediate needs. What citizens want in information from testing programs, for example, and what we want can differ. What does not differ is the common goal of providing the best possible education for the youth of our society. And, as suggested in most of these articles, there is no reason why well-designed testing programs cannot provide information to serve the needs of educators and our patrons. The challenge is clear for readers: Do the testing practices followed in your school meet the information needs of educators and the community? Are test results communicated to those with a need to know? Those with a right to know? Are test results used for the benefit of individual students, for the improvement of instruction, and for the benefit of the community?

An End to the Debate. If it has not already ended, perhaps this issue can be a "catalyst" to end the debate of the past two decades over the relative merit of norm-referenced and criterion-referenced assessment of learning outcomes. Both types of data are needed. By definition, a norm exists whether we have assessed it or not and, indeed, normative behaviors and patterns of behavior are the basis for a soceity or, in the academic world, a discipline or a professsion. At the same time, part of our cultural heritage and our educational

heritage is a concern with the rights, responsibilities, and performance of the individual. As should be true in more areas of life, the operational word is "and" rather than "or."

A Look Ahead. The next issue continues a theme of "partnership" between education and its clientele by examining changing conditions in family life and the implications of those changes for educational policy and practice. During the past decade, changes in the economy and changes in patterns of involvement in the labor force have caused significant changes in the traditional patterns of child-rearing which are held as ideals by many Americans. These changes have been further extended by changes in life styles and changes in values, including major changes in the value which is placed on having or rearing children. These trends are examined in depth in the next issue.

Reminders. We want your ideas, your feedback, your articles, your suggestions for topics and authors. Remember that back issues of **Catalyst** can be ordered from the NCSA offices; see the Fall, 1980 issue for details.

Edgar A. Kelley
Winter, 1981

take a good look

at the only microcomputers designed exclusively for educational use

ONLY CENTURION PRE-PROGRAMMED COMPUTERS ARE DEDICATED EXCLUSIVELY FOR EDUCATIONAL APPLICATIONS!

American education needs the benefits available from microcomputer technology in basic skill subjects such as arithmetic and reading. Centurion's pre-programmed "dedicated" computers support specific learning functions in well-defined curriculum areas. Coupled with an outstandingly simple, easy-to-administer Management System, Centurion microcomputers offer highly motivated individualized learning in basic arithmetic and reading skills.

THE DIGITOR® LEARNING ARITHMETIC MODULE

Introduced to education in 1974, the Centurion Digitor microcomputer is now successfully supporting basic arithmetic drill and practice in over 50,000 classrooms throughout America. With Digitor, students individually achieve MASTERY level learning of computation skills in addition, subtraction, multiplication and division.

Digitor is available in three individual microcomputer models:

The Tutor. The basic learning model for students from Kindergarten up, incorporating all 0-9 number facts in groups of 10, 25, 50 and 100 problems with elapsed time controls.

The Drillmaster. The basic arithmetic drill model, incorporating all 0-9 number facts with 10 problems per series and elapsed time controls for power drill exercises.

The Skillmaster. The MASTERY level learning model, incorporating all 0-9 number facts with semi-automatic and fully automatic response time control and random mixing of arithmetic operations.

THE ALPHATOR® WORD ATTACK MASTERY MODULE

The Alphator is Centurion's newest microcomputer for student basic skill MASTERY level learning of the basic word attack skills required for competency in reading from grades 1 through 5 and above.

The Alphator Learning System incorporates **35** individual practice lessons in the nine subject areas of **Letter Matching, Vowel Sounds, Consonant Blends, Compounds, Affixes, Syllables, Spelling, Ordering** and **Word Review**. It contains **177,987** randomly presented individual questions for student response. Therefore, students will be continuously enthused and motivated to practice skills until individual MASTERY is achieved.

Alphator is available in three individual microcomputer models:

The Alphator I, containing a total of **16** lessons in four subjects utilizing a vocabulary of **441** words.

The Alphator II, containing a total of **16** lessons in four subjects utilizing a vocabulary of **562** words.

The Alphamaster, containing a total of **35** lessons in nine subjects utilizing a vocabulary of **701** words in **177,987** individually presented questions.

Centurion computer-based learning systems are most cost effective for providing highly motivated practice and drill. Centurion's dedicated computers are priced from $119.50 to $399.50 each. They require no additional programming time or expense. Savings realized in student workbooks alone can make the cost of Centurion Learning Systems self-liquidating within the first school year of implementation.

NO COMPUTER SKILLS REQUIRED

Of special importance is that Centurion Computer Learning Systems require no teacher skill in computer programming. Centurion computers have their learning programs already included in the computer memory to effectively teach MASTERY level learning in each subject. Further, Centurion Learning Systems use simplified record-keeping methods for student progress that correlate directly with any required competency testing.

THE ONLY MICROCOMPUTERS WITH A PROVEN RECORD OF EFFECTIVENESS IN THOUSANDS OF CLASSROOMS

Research studies over the past 5 years prove Centurion Learning Systems support MASTERY learning to the highest degree while saving teachers' time and truly motivating students to learn.

Free classroom evaluation available: Contact your local authorized Centurion dealer or Centurion for a no-obligation classroom evaluation or write for a free full-color brochure.

centurion
industries incorporated
167 Constitution Drive, Menlo Park, California 94025

Stephenson School Supply Co.
P. O. Box 81207
Lincoln, Nebraska 68501

GUEST EDITORIAL

L. JAMES WALTER

TOWARD IMPROVED TESTING PROGRAMS

Regular testing of large numbers of students is a practice which has been faithfully followed by school personnel for many decades. During most of that period of time, testing practices were not a cause of controversy. School officials seldom used the results of the testing program as a basis for significant program changes and the existence of widespread testing meant, for the most part, that a variety of scores were recorded on the permanent record of each student. These scores were used to explain student classroom performance and, on occasion, as a basis for placement of students in courses.

As the media began to report a pattern of declining test scores, particularly scores on the Scholastic Aptitude Test (SAT), school patrons voiced increasing concern about the quality of schools. In the past few years, this concern has been made manifest in a number of ways. Competency testing programs for both students and teachers, calls for a return to the "basics," and significant reductions in curricular offerings (especially the elective curriculum) are only a few of the responses which originated with the public's concern over declining test scores.

School officials, faced with public concern over the quality of schools and schooling, tried to defend the curricular and instructional practices already present in schools and offered a number of explanations as to why a pattern of declining test scores had developed. Examination, both of the criticisms offered by patrons (and, at times, educators) as well as the statements offered in defense of schools and schooling, leads to a number of general conclusions about the state of the art in the use of tests and testing programs:

1. In most school districts, there is a lack of a clearly stated philosophy or rationale for either the testing program which is being followed or the one which the school and community wishes to have and use.
2. In many school districts, there is a weak or limited relationship between the content being taught as a part of the cur-

L. James Walter is Associate Professor, Center for Curriculum and Instruction, University of Nebraska-Lincoln and Guest Editor of this issue of **Catalyst.**

riculum and the content being tested by the tools used in the testing program. In all but a handful of districts, the relationship between the curriculum and the testing program is examined infrequently if at all.

3. Many educators lack the skills necessary for use of test data as a basis of planning for curricular and instructional decision-making and, at the same time, are unable to adequately explain the meaning of test data to students, parents, or patrons.
4. While test data frequently are described by educators as being necessary so that instruction can be individualized for students, this worthwhile goal is honored as much in the breach as in practice.
5. As the public has become more concerned about and interested in test results, educators have tended, to an increasing degree, to take credit for favorable test results and to seek to explain why unfavorable test results are not important. These defensive behaviors have not fostered increased confidence among the public in their views about the quality of schools and schooling.

While other concerns are also present, the five conclusions which are listed appear to be the major issues which should be addressed by educators in each school and school district. The shortcomings listed in these five conclusions can be overcome. Some specific recommendations for action include those which are listed below.

1. A clear statement should identify the philosophy or rationale of a school district's testing program. If such a statement exists, it should be followed in practice. If such a statement does not exist, such a statement should be formulated with widespread participation by both faculty and patrons of the school district. Discussion of a testing philosophy or a rationale for testing programs and practices should center on the uses to be made of test results as well as determination of appropriate times at which tests should be administered. Recommendations regarding testing programs and practices should be compatible with the overall school philosophy.
2. Tests which are used should be consistent with curriculum and instruction being provided in the school district. The best match between testing practices and instructional practices usually can be found when criterion-referenced tests are tailor-made for the curriculum which is offered. Such tests often are expensive. If a standardized achievement test is to be selected for use, as would be a more common practice

than development of a criterion-referenced test, educators in the school district should be certain that a content analysis of the test and the district's curriculum is conducted so areas of consistency and inconsistency are identified. Such an analysis before test data are collected can help to improve understanding and use of data as well as improving the ability of educators to explain test results to patrons. (Of course, being able to do this assumes that curriculuar and instructional planning is clear and concise within the district.)

3. School district personnel need to understand the tests which are administered and the data which are available for use. The inservice or staff development programs of the district should invest necessary resources in being certain that staff understand the testing program and can use the data provided. Using educators within the district who are already skilled in tests and measurements or training a small number of existing personnel, can offer a pool of resource people within the district who can be used to train other district personnel.
4. If the philosophy which is adopted accepts the traditional premise that test results should be used to make adjustments in instruction being provided to students, the district should plan to make specific test data available to teachers and should expect that teachers and building administrators will use such data to make adjustments in instruction or adjustments in the placement of students. To follow such practices will require frequent testing, opportunities for retesting, and the availability of instructional alternatives.
5. Educators need to be objective in the use and interpretation of test scores. Most educators would agree that schools, alone, can not be blamed, totally, when test scores fall or are low. Hopefully, educators also would agree that schools, alone, can not take all of the credit when student test scores improve or are high. Having a better understanding of what test scores can and can not tell educators and patrons about student performance and the quality of schools can help to provide the objectivity which educators need to display.

As a closing note, school districts should plan to review their testing philosophies and practices on a regular basis; a recommendation made elsewhere and repeated here is that such a review should occur every three to five years.

Raymond G. Jensen
4127 Garfield Street
Lincoln, Nebraska 68506
Phone: (402) 488-4464

Marcel A. Devos
1404 Bellevue Blvd. N
Bellevue, Nebraska 68005
Phone: (402) 731-5189
Whse: 1015 No. 14 St.
Omaha, Nebr. 68102
Phone: (402) 342-4248

Arthur P. Williams
Whse: 822 West 16 Street
Hastings, Nebraska 68901
Phone: (402) 463-7131

"Sanitation with Service"

165 Combined Years Of Experience To Serve You

Tim D. Shea
6405 South 76 Circle
Omaha, Nebraska 68127
Phone: (402) 331-4486
Whse: 1015 North 14 Street
Phone: (402) 342-4248

John H. Perkins
1015 North 14 Street
Omaha, Nebraska 68102
Phone: (402) 342-4248

Marcel A. "Skip" Devos
201 Sycamore Circle
Gretna, Nebraska 68028
Phone: (402) 332-4615
Whse: 1015 North 14 Street
Omaha, Nebraska 68102
Phone: (402) 342-4248

Customer Satisfaction Guaranteed!

Disinfectants and Deodorizers • • • D/S/D — Lila-Phene-Odorless "A"
Carpet and Wall Spotters and Cleaners • • • Break-Thru — Spra-Kleen — Dri-Est Foam
Dust Allayers • • • G.F. Maintainer — Dustphene — Lemon Oil Polish
General Cleaners • • • Hi-Con — S-33 — Klenzall
Floor Cleaners • • • Mist 'N Kleen — Mist 'N Polish — Dual Treet
Floor Finishes • • • Super Market Finish — Super Wear Cote — Nu-Cote

Call (Collect) (402) 342-4248

Industrial

Chemical Laboratories, INC.

1015 NORTH FOURTEENTH STREET • OMAHA, NEBRASKA • 68102

CONRAD CARLBERG

In 1980, the Mid-continent Regional Educational Laboratory (McREL) completed a survey of testing practices in Nebraska. The findings of that survey, and their implications, are discussed in this article. Among the suggestions are strategies for using data from existing testing programs even when the test battery is not highly congruent with the curricular and instructional program.

TESTING PROGRAMS IN NEBRASKA SCHOOL DISTRICTS

In early 1980, the Nebraska Department of Education asked McREL* to survey testing policies and procedures in Nebraska school districts in order to learn more about the kinds of assessment issues that might be of concern statewide. This article describes some of the general findings. Specifics, such as the survey instruments used, the sampling design, the survey methods and findings not reported here—and of interest primarily to test specialists—are available from both the Nebraska Department of Education and from McREL. (These items are beyond the scope of this article. Also, they're boring.)

Twenty Class II and Class III school districts and 112 superintendents, principals, testing coordinators and teachers were surveyed. Questions about test selection, purposes for testing, use of test results, and feelings of frustration or success with testing programs were asked.

Findings About Test Purposes

School staff identified three basic groups of purposes for testing. They assigned greatest importance to purposes of individual student assessment, i.e., diagnosis of academic strengths and weaknesses for individual students, identification of student growth and progress, and the identification of students needing special instruction.

The next most important set of purposes were those that serve guidance and program evaluation functions. These purposes in-

Conrad Carlberg is Director of Research, Mid-continent Regional Educational Laboratory, Denver, Colorado.

*McREL is one of the national network of regional Labs and Centers. It provides research-based services to states and districts in its region which includes Nebraska, Colorado, Kansas, Missouri, North Dakota, South Dakota and Wyoming.

cluded curriculum evaluation or change as well as assistance to students in making decisions about courses of study.

Purposes of external comparisons were seen as least important. These purposes include comparison of district test scores with national norms and comparisons between and among attendance centers.

Findings About Test Selection Practices

The two most frequently mentioned reasons for selecting particular tests or batteries of tests were tradition and usability. Cited less frequently were test content, test reliability or test validity. Other reasons (e.g., evaluation or availability of scoring services) were rarely mentioned.

Tradition can be a good reason for selecting a test. It makes possible the comparisons of scores over several years, and it eases the problem of score interpretation.

Usability subsumes questions of economy, time constraints and interpretability. The importance of each is evident. Educational funds are scarce (economy); the time spent in testing detracts in part from time that can be spent on instruction (time constraints); and test scores should not exist in a vacuum, but should be interpreted by reference to educational objectives (interpretability).

Findings About Feelings of Frustration

School personnel were asked to identify the most frustrating aspects of their testing programs. More than half of the respondents were frustrated by the lack of use of test results. The next most frequent responses were students' frivolous approach to tests and the cost of testing.

The lack of use of test results as a cause of frustration comes into sharper focus when questions of test purpose and test selection are considered. The three most important purposes identified were diagnosis of individual students' strengths and weaknesses, measurement of individual student growth, and identification of special needs students. If test scores are to be used for these purposes, content relevance and high reliability are essential test characteristics. But, apparently, these considerations have not been applied to the selction of tests as much as have considerations of tradition and usability. It is not surprising, then, that staff are frustrated by the lack of use of test data.

Even more specifically, the two tests most commonly used in Nebraska school districts—the Science Research Associates (SRA) and Iowa Tests of Educational Development (ITED) batteries—do not address themselves to specific course objectives. The content

of these measures stresses the accommodation of as broad a range of curriculum and of students as possible; in fact, the ITED stresses cognitive processes and de-emphasizes content as much as possible. the general reliability of these tests is based on a more heterogeneous population than exists in individual Nebraska school districts. This means that the ability of standardized test results to identify individual student differences is limited. Therefore, while standardized tests such as SRA and ITED are admirably suited to functions such as external district comparisons and, to some degree, program evaluation, they are less relevant to individual student assessment. The frustration over the lack of use of test scores is virtually inevitable.

Implications

An exceptional degree of consensus among respondents on the relative importance of different purposes for testing was found. Administrators (i.e., superintendents, principals and counselors) all felt that the most important purpose of testing is to identify the strengths and weaknesses of individual students. Teachers ranked this purpose as second only to monitoring student growth.

These findings argue strongly that any testing decisions made by school districts should stress the area of student assessment heavily. To draw the distinction between assessment decisions and other decisions more clearly, the former should be made on criteria such as: "Is this test useful for locating students with low language arts skills and for monitoring their progress?" A less effective decision criterion would be: "Is this test useful for evaluating a program that enhances students' language arts skills?" This latter decision criterion could be met by means of tests whose reliabilities (and, possibly, whose content validities) are only moderate, because the purpose is to discriminate among groups; but to make instructional and assessment decisions about individuals, high reliabiljty and strong content validity are both absolutely essential.

Findings on the considerations that are employed in choosing a test reflect both theoretic and pragmatic concerns. Current psychometric thought suggests that the debate over criterion-referenced or norm-referenced testing is following the classic nature-nurture path: we are beginning to recognize that both kinds of interpretation of test scores are important. This has probably always been recognized at the local level; the teachers in our sample, for instance, were primarily responsible for what emphasis there is in the overall findings on test content (pertinent to criterion-referenced testing). Teachers also find reliability and validity concerns of importance, although less critical than test content. This

would argue that teachers are primarily interested in the match between the curriculum and the test (for instructional, criterion-related purposes) and secondarily concerned with the psychometric aspects of tests that allow more precise identification of individual students' strengths and weaknesses.

Administrators and counselors, while concerned with test content, appear to be more concerned with the technical and pragmatic aspects of tests. Perhaps because they often must take a school or district-wide view, they want (a) to see test information put to the best use possible (usability, interpretability), (b) testing and score interpretation to proceed as smoothly as possible (ease of administration, tradition, and availability of various scoring options), (c) to be able to establish relationships among test scores and other variables such as differences among programs and among institutions (e.g., district-state; district-nation; school-state; school-school; etc.), and (d) to identify growth, strengths, and weaknesses for individuals (reliability and validity).

It has been demonstrated (Floden **et al.**, 1978; and others) that the content covered by various popular standardized achievement tests can vary substantially from test to test. That this is so is illustrated in Figure 1.

Figure 1: Percent of Items Measuring Different Mathematics Operations on Four Standardized Tests.

Operation	ITBS	MAT	SAT	CTBS
Add	13%	21%	13%	14%
Subtract without borrowing	8%	8%	6%	8%
Subtract with borrowing	11%	11%	6%	5%
Add and subtract fractions without common denominater	1%	0%	0%	2%
Multiply	11%	19%	16%	17%
Divide without remainder	6%	9%	15%	14%
Divide with remainder	1%	1%	1%	1%
Combination	8%	6%	7%	7%
Grouping	2%	0%	5%	0%
Identify equivalents	20%	18%	16%	15%
Identify rule	11%	3%	9%	12%
Identify terms	8%	5%	6%	4%

Adapted from Floden **et al.**, 1978. Entries are percent of items for fourth grade mathematics. Tests are: Iowa Tests of Basic Skills (ITBS), Stanford Tests of Basic Skills (SAT), Metropolitan Achievement Tests (MAT) and the CTB/McGraw-Hill Comprehensive Tests of Basic Skills (CTBS).

Such a situation is not problematic when a purpose is to predict a student's future success, or to assess the student's knowledge relative to other students. But, as noted above, these are not primary concerns to the Nebraska educators in our sample. Their concerns are to identify individual strengths, weaknesses and growth. For instructional purposes, these concerns are relative not to some norm group, nor to some performance criterion, but are instead relative to the student's overall level of functioning and in specific subject-matter areas.

It is for this reason that test choice must include test content as a decision criterion. If, for example, a district's fourth grade mathematics curriculum stresses "subtraction without borrowing" more than "division without remainder," that district's teachers will be rather disappointed with the mathematics results that their students achieve on the SAT (6% subtraction, 15% division); while another district with a similar fourth grade curriculum will be charmed by its mathematics scores on the ITBS (11% subtraction, 6% division). Similar students and similar curricula will be assessed differently, because the tests were not chosen with due regard for test content.

Recommendations

The foregoing does not imply that McREL recommends either tailoring the curriculum to test publishers' emphasis; or switching tests solely because another publisher's test empnases are more congruent with the district's curriculum; there are other criteria that must be met for a major switch to be desirable. But it is possible to interpret scores in light of the degree of test-curriculum match: "testing to teaching." Prior to receipt of scores, teachers can specify their approximate instructional emphasis in various item categories and subsequently evaluate students' scores on this basis. This requires that there be access to item-level norms and to the actual items administered.

Because teachers inevitably, and often rightly, put more faith in tests they make themselves than in tests made by others, teachers are inclined to use results of their own tests more fully than the results of standardized tests. Because thoughtfully prepared teacher-made tests reflect a classroom's instructional content more closely than any commercially available test possibly can, McREL recommends that assistance be provided to teachers in classroom test construction and interpretation. Good classroom tests, when accumulated, can attain reliabilities in excess of those that are characteristic of standardized tests; they can reflect curricular emphasis much more accurately; they can yield valid in-

ferences about individual students; and they can provide criterion-referenced interpretations. If they are constructed from some sort of statewide or districtwide item bank, they can provide norm-referenced interpretations; however, they need not do so if the standardized test is well chosen, and interpreted with care.

1. Floden, R.E., Porter, A.C., Schmidt, W.H., and Freeman, D.J. *Don't they all measure the same thing? Consequences of selecting standardized tests.* Research Series No. 25, Institute for Research on Teaching, Michigan State University, 1978.

SCHOOL SPECIALTY . . .

FOR CONVENIENT ORDERING OF SUPPLIES AND EQUIPMENT!

* Toll-Free phone number: 800-255-0130
* Free catalog available upon request
* Large stock of popular forms, supplies and materials
* Area representatives for personal service:

John Hauptli
204 W. 23rd St.
Grand Island, NE 68801

Gary Runyan
Box 1327
Salina, KS 67401

John Pollard
Box 404
Loveland, CO 80537

School Specialty Supply, Inc.

P.O. Box 1327
Salina, Kansas 67401

3525 South 9th Street
Phone 913-827-0451

Call Toll-Free: 800-255-0130

H. JAMES FUNK
CHARLES R. DuVALL

All school districts, to one degree or another, have felt the pressures of state and national movements toward competency testing and increased accountability as well as the pressures which have arisen from public concern over declining test scores and public demands for emphasis on basic education. To the extent that these pressures help educators and school districts examine their testing programs and their approaches to curriculum and instruction, a number of benefits are likely. In responding to these pressures, many school districts will continue to rely on commercial tests as the bulwark of district testing programs. The authors describe some of the questions which educators should examine when considering the adoption of commercial tests.

SO, YOU'RE GOING TO BUY ANOTHER STANDARDIZED TEST... WHOSE? WHY ?? WHO CARES???

Like it or not, schools are getting on the accountability bandwagon! The furor accompanying testing and accountability is healthy for school systems, professional staffs and students. It provides the basis for a dialogue concerning something about which we all ought to be concerned (achievement) and something about which everyone can argue (results). It can lead to more testing (a possible good), a closer look at what students learn from the curriculum (a definite good), and a closer look at what standardized tests really measure (another definite good). Still, many persons undoubtedly will continue to opt for choices of tests based upon a salesperson's presentation, cost, or what the surrounding school districts are doing, rather than if the test measures anything of consequence or whether the test's contents (questions) have any relationship to the curriculum. For those choosing tests for any of these reasons, STOP READING! Simply buy the test and prepare for the dismal, expectable results, i.e., students, teachers, and the school district will fare poorly. The time saved by not reading this article can be used in preparing a good list of excuses. For the more hardy, read on!

Why Test?

One reason for having a testing program is that the public

H. James Funk is Associate Professor of Education and Charles R. DuVall is Professor of Education, Indiana University at South Bend, South Bend, Indiana.

demands it. This is a very good reason. No school exists independent from the constituency that supports it and the public wants to know how their schools are doing. Parents also want to know how their children are doing in school. The community may view a program of standardized testing as an objective, external measure of control over the quality of schooling.

A second reason for testing is to improve the quality of instruction. Tests can provide valuable information regarding the strengths and weaknesses of a curriculum. Test data can also aid in the placement of students in appropriate areas of instruction and remediation or be used for comparison of student growth across areas of instruction.

Test Validity

The single most important feature of any test is content validity. Simply stated, this means the ability of a test to measure what is being learned. The best way to determine content validity is to provide staff members with sample copies of the test and ask them to analyze the test, question by question, asking themselves: "Are these questions testing what I am teaching?" The test (or battery of tests) having the highest number of individual items best measuring a particular curriculum will probably be the better test for that school. By involving all teaching staff, it is possible to avoid the trap that many school administrators fall into—selecting a test that does not measure what is being taught. Teachers are interested in test results, particularly if they believe the test measures what is being taught. Involvement of staff in test selection has other possible benefits; it can lead to a closer examination of the school's curriculum and it may reveal areas in which there is a lack of congruence between the objectives of the school and teaching practices.

Working With Salespersons

In working with the salesperson, avoid the question: "How good is your test?" All salespeople believe their product is superior and that it can, under optimal conditions, predict the mean daily temperature in David City, Nebraska. Despite this penchant for exaggeration and hyperbole, most salespersons can give useful information regarding the quality and use of the tests they sell. Some basic questions to ask are:

1. How was the test validated?
2. How are the scores reported?
3. What norm groups are available?
4. What scoring and reporting services are available?
5. What are the costs involved in use of the test or test battery?

It is not the responsibility of a test publisher to make a test valid for a particular school district; however, the publisher producing a more valid test should sell more tests. How were the tests validated at a general level? What procedures are used in other schools to validate the test? The salesperson should be able to provide names, addresses, and contact persons for independent inquiry if such queries are part of a school district's review practices.

In asking questions about norms and scores reported, it is important to know whether the test provides norm-referenced data, criterion-referenced data, or both kinds of data. It is equally important to know whether or not the populations to whom the test has been administered are comparable to the population of the school considering its use. In general, norm-referenced tests are used to compare the growth of one student population with another while criterion-referenced tests are used to determine student attainment of specific objectives. Can norm-referenced tests be used to compare individual student growth as well as group progress? Yes, but only if the norm-referenced test (which lacks a specific set of objectives) is turned into a criterion-referenced test through the use of individual item analysis (matching test questions to curricular objectives) by the people actually doing the teaching. Can criterion-referenced tests be used to compare groups of student populations? Yes, but only if the test was normed with a comparable population and the objectives measured by the test are the objectives of the school in which the test will be used. Before a test is selected, answers to these questions should be established through staff analysis of the test or test battery.

Careful examination of the services which a test publisher can offer and is willing to offer is warranted. The availability, selection and use of appropriate services can make the difference between testing for the sake of testing (a fruitless exercise) and testing to benefit students and the instructional program of the school. It can be assumed that all companies will provide data such as group comparisons and scores. Further, the publisher should offer a scoring service (the authors cannot imagine any administrator asking a staff to hand-score the tests). Be certain data are present in appropriate reporting formats—ones that fit the needs of the school district. A key question is "How will information obtained be used?" Once the uses of the data are determined, the publisher can be asked to present the information in a manner compatible with a school's information needs. Additional services to be considered are analyses of student performance in individual subject or skill areas. Some companies provide printouts of individual student performance with respect to student strengths and weaknesses on objectives measured by the test. This type of service can be of maximum value

only if inservice training is provided for teachers on interpretation of data and on ways to communicate test data to parents. If the school district does not have personnel resources to provide such inservice, the salesperson may be able to help arrange for speakers at inservice sessions or sessions with parent and patron groups. Make sure the persons selected to make presentations can communicate in language which can be understood by the groups to which presentations are being made.

Review of Sample Tests

What criteria should be used to review sample tests? Publishers usually provide specimen sets of tests at little or no cost. Individual tests, rather than an entire battery, should be examined. It is possible, even probable, that a test may be excellent in one or two curriculum content areas and weak in others. For example, there are many good tests for basic skills areas but when an entire battery is considered the areas of social studies and science often are weak in treatment. It is probable that this can be attributed to an emerging consensus about desired curriculum content in reading and mathematics; the same degree of consensus is not present for science and social studies curricula.

Test selection processes are complicated by the fact that the potential for what tests can measure increases as students get older, i.e., tests become more complex and diverse at upper grade levels compared to lower grade levels. With increased diversity, the potential for mismatching curricular objectives and the testing program also increase. These differences make it more difficult to choose a single publisher's series of tests (battery) as an entire, and appropriate, testing program for a school district.

Wise Use of Test Results

Test results can be used for appropriate or inappropriate purposes. Serious questions must be raised about whether or not the public's "right to know" includes wholesale dissemination of test data when such data are used to make comparisons between schools and among teachers. Public disclosure of group test data may increase newspaper sales, help real estate agent sell homes in "selected areas," and also help test publishers sell more tests. Such uses of test data are not recommended.

Intelligent dissemination of test scores to parents and teachers, for the benefit of individual students and for the improvement of programs of curriculum and instruction, is an appropriate use. Schools can benefit from honest and accurate appraisals of academic progress in curricular programs. Teachers benefit by use of specific information regarding the strengths and weaknesses of

their students' performance and from the opportunity to adjust instruction to the needs of students. Students benefit from instruction which is designed to build upon existing strengths and correct existing weaknesses. Parents benefit from receipt of accurate reports about their children's academic progress. And, the community benefits when it feels assured that the school is working in a systematic manner to improve instruction.

Summary

Standardized testing practices, with regard to what tests measure and to how test data are used, should be improved. Persons involved in the administration of tests or in the use of test data must have a clear understanding of the purposes of the testing program and should be involved in the selection or design of the testing program. The instrument or instruments selected should relate to the school's curriculum and should test what actually is being taught. Data provided by a testing program should be reported in such a fashion that relationships to the instructional program are understood and accepted by the staff, students, parents and community. Valid tests, properly administered and interpreted, can be immensely beneficial in the improvement of instruction—and instruction is what schools are all about!

YOUR CHILDREN'S
SAFETY
IS OUR BUSINESS

BLUE BIRD School and Transit Buses
SALES and SERVICE

MISLE BUS & EQUIPMENT CO.
50th & O Streets Lincoln, Nebraska 68510
Phone: (402) 483-2261
Julius Misle Paul Jablonski
Call Us TOLL FREE for All Your Needs: 800-742-7605

"ORIGINATORS OF ALUMINUM SEATING"

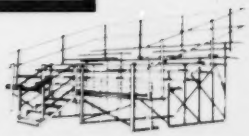
we're

stadiums unlimited, inc.

still no. 1

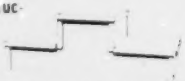
GRANDSTANDS

• Alumn-a-Plank Portable Bleachers • Alumn-a-Plank Tip & Roll Bleachers • Indoor Roll-a-way Wood Seat Alumn-a-Plank Replacement • Benches - Team



and / or Park Design • Mini - VIP Reserve Seating • Portable Judges Stands • Awards Stands • Alumn-a-Plank Renovation • Alumn-a-Cover

for Wood Seats • TOT-DOCK - Underwater Instructional Swimming Platform • Picnic Tables



For further information write or call...
Free color catalog on request...

"RAPiD ROBERT"

"America's No. 1 Bleacher Salesman"

Robert H. Gilbert,
Vice President, Sales
3516 East 8th Street
Des Moines, Iowa 50316
Phone 515/266-9694

PAUL G. MONKOWSKI

To improve instruction and to respond to the needs of individual students, educators need criterion-referenced test data. The public, however, has an interest in norm-referenced test results. Both information needs must be served by most testing programs. In this article, the author describes and illustrates how expectations of both educators and the public can be met.

PRACTICAL USES AND INTERPRETATIONS OF ACHIEVEMENT TESTING

Advances in computer technology during the 1970's have created a revolution in academic testing for school systems. Prior to these advances in technology, testing was primarily seen as a nonfunctional activity from the perspective of classroom teachers and also many school administrators. Information gleaned from testing provided minimal assistance to teachers in identifying strengths and weaknesses of individual students other than in a global sense, i.e., this student performed better in Reading Vocabulary than in Reading Comprehension. Hence, testing was seen as an administrative hassle with the benefits being derived by some administrator (usually the district's token statistician) who used the data for reports to the Board of Education and various government agencies. Curriculum specialists also found minimal application of test information to aid them in their work primarily for the same reasons. Information provided tended to be interpreted in an absolute sense without any consideration given to chance level fluctuations in scores or what is known in psychometric terms as the standard error of measurement.

In order to supply their need for individual diagnostic information which would then permit a more indepth analysis of the strengths and weaknesses of the school curriculum, classroom teachers and curriculum specialists teamed up to supplant norm-referenced testing with criterion-referenced testing. CRT was seen as the future of testing; it could be locally developed to reflect the curriculum thereby permitting curriculum specialists to monitor the curriculum, and it was specific enough so as to provide meaningful diagnostic information at the classroom level to aid the teacher in more adequate skills grouping, and identification of strengths or

Paul G. Monkowski is a doctoral student at the University of Nebraska-Lincoln. His work experience includes employment by the University of Nebraska-Lincoln, the Lincoln Public Schools and McGraw-Hill in the area of educational tests and measurements.

weaknesses. It appeared that the CRT movement was the answer to the frustration of classroom teachers and curriculum specialists. Yet, some very real problems in this movement were overlooked.

CRT's as originally designed had no data to support decisions made by academic personnel regarding student performance and curricular trends because the psychometric properties of CRT development were in their infancy. Hence, locally developed CRT's had, at best, attended to face validity, i.e., the local developers verified the validity of the test items and criteria for determining mastery of the objective. Likewise, end-of-chapter tests published along with basal series had (and continue to have) limited empirical data to support their interpretation. Criteria for mastery were determined by simply counting the number of items correct out of the total number of items measured by the objective without taking any other variable into consideration such as the difficulty of the items missed and of those correct, the relationship between performance on the objective being measured and the performance on other objectives of the test, and the relationship of the items of the test to each other. (Taking these factors into consideration, however, can serve to increase the accuracy of the information which in turn reduces judgmental errors in interpreting test scores and decisions associated with test scores.)

A primary difficulty with CRT's was their inability to provide norm-referenced data regarding student performance. Boards of Education and parents objected to reports of "number of objectives mastered and not mastered" in content areas and were insistent in their demands to know the relationship between individuals' achievement relative to other students, i.e., are they at the same level as others in that grade? The public demands norm-referenced data, while the teaching personnel has much greater use for CRT information. The two approaches needed to be linked in order to satisfy the needs of both educators and the public.

The California Achievement Tests (1977 edition) were the result of the first merger of norm-referenced and criterion-referenced approaches to measurement of school achievement. These tests were developed from objectives commonly found in basal texts and state department curriculum guides. Items were written to measure specific objectives in addition to contributing to the overall test score. This resulted in both norm-referenced and a form of criterion-referenced data being made available from a single test administration of an empirically valid and reliable instrument. The legimate concerns of educators in the past regarding the usefulness of test information therefore have been attended to and can no longer be considered valid reasons for ignoring test information.

The increase in the sophistication of information provided by

academic testing has necessitated staff development in test interpretation and utilization. Both classroom teachers and administrators have a responsibility to their constitutents to properly understand principles of testing and to appropriately interpret test results. This does not entail memorization of statistical formulas and psychometric terminology, but it does require a willingness to learn the concepts vital to correct test use and interpretation. What follows is an interpretation of an individual student record of performance on the CAT with suggested ways for utilizing this information for instruction by the classroom teacher and an analysis of district performance to aid the curriculum specialist in monitoring and updating curricular needs.

DERIVED OBJECTIVE MASTERY
Mastery Performance

Reading Vocabulary	
Obj. 32 Same Meaning	-
Obj. 33 Opposite Meaning	+
Obj. 34 Multimeaning	P
Reading Comprehension	
Obj. 36 Recall of Facts	+
Obj. 38 Inferred Meaning	-
Obj. 39 Character Analysis	-
Obj. 40 Figurative Language	-
Obj. 42 Author/Attitude Position	-
Obj. 43 Techniques of Persuasian	-

▪
•
•

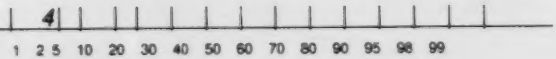
The Individual Student

Table 1 is a display of part of an Individual Test Record (ITR) from an administration of the California Achievement Test for a hypothetical sixth grade student. Examination of the top portion of the ITR, reveals the usual norm-referenced information regarding performance on the overall battery and its various tests. Of interest in the top portion is the graphic representation of the national percentile bands for the tests. The percentile band represents the range in which the student's true score is likely to fall. This range represents a reasonable fluctuation in performance due to factors beyond the control of the testing situation. This "reasonable fluctuation" is the result of applying the psychometric concept of standard error of measurement to test performance, thereby moving in-

Table 1

Individual Test Record for a Hypothetical
Sixth Grade Student

TEST	Raw Score	Grade Equivalent	National Percentile	Percentile Band
Reading Vocabulary	22	6.2	42	XXXXX
Reading Comprehension	23	5.2	33	XXXXX
Total Reading	45	5.7	37	XXX
•				
•				
•				



terpretation away from seeing performance in an absolute sense, i.e., the student's score corresponds to the 42nd percentile in Reading Vocabulary; and toward seeing performance in a relative sense, i.e. the student's score in Reading Vocabulary corresponds to scores somewhere between about the 39th percentile and the 52nd percentile.

The lower portion of the ITR displays mastery scores for the objectives that compose each of the tests listed in the upper portion. For example, the Reading Vocabulary test is composed of items that measure the three objectives found in the lower portion of the ITR: Same Meaning, Opposite Meaning and Multimeaning. The student's overall performance in Reading Vocabulary placed the student in the 42nd percentile nationally but within that test the student demonstrated mastery of the objective of Opposite Meaning (+), partial mastery of Multimeaning (P), and nonmastery of Same Meaning (-). If the classroom teacher wishes to provide instruction in Reading Vocabulary, the teacher can best meet this student's individual needs by emphasizing same meaning vocabulary activities since this is the identified area of greatest need. This lower portion of the ITR povides the classroom teacher with information that is specific enough to be sensitive to instructional planning and activity. The normative data reported by the top portion of the ITR provides the teacher with global information regarding individual strengths and weaknesses. Once these areas have been identified, the lower portion provides more specific information regarding performance on objectives that make up the test. Thus, the ITR becomes a useful tool in diagnosing individual needs and developing educational intervention strategies to meet those needs without sacrificing norm-referenced data.

District Performance

The summary of overall district performance can provide information of benefit to administrators and curriculum specialists in addition to classroom teachers. Table 2 summarizes district level performance on a portion of the California Achievement Test. Performance on the objectives of the test can be aggregated to produce summary results for the district regarding the relative strengths and weaknesses of the district. From this information, priority objectives can be isolated for further curricular development and refinement.

Table 2 shows the difference in percent correct between the national norm group and the local district for the objectives that compose the Reading Comprehension test. A value of zero in this "percent correct difference" column indicates that the local district per-

Table 2
Right Response Summary

	Percent Correct Difference Between Local and National	Priority Rating
READING COMREPHENSION	-2	
Literal		
Obj. 36 Recall of Facts	-4	2
Interpretative		
Obj. 38 Inferred Meaning	0	4
Obj. 39 Character Analysis	-2	3
Obj. 40 Figurative Language	-12	1

formed on par with the national norm group. Consequently, a negative value would indicate poorer performance at the local level than the performance of the norm group, and a positive value would indicate a better performance at the local level than the performance of the norm group. Relative strengths and weaknesses in the curriculum can then be identified by determining the degree to which performance by the district on each objective differs from zero, the reference point of the norm group. The "Priority Rating" column of Table 2 shows that the local district had the greatest difference in performance when compared to the norm group on Objective 40, "Figurative Language" followed by Objective 36, "Recall of Facts."

Table 3 continues the process of identifying district strengths and weaknesses by breaking down Objective 40, "Figurative Language", already identified as a priority objective from analysis of Table 2, into its item composition and applying the same principle as outlined above. This item breakdown reveals that student performance is below the national norm group on items dealing with simile.

This systematic breakdown of district performance on objectives and items allows the content specialist to pinpoint specific areas in the curriculum in need of improvement; in this case, simile. This process, therefore, leads to a systematic method of allocating resources for curricular improvement in areas where the resources are most needed.

In conclusion, the successful implementation of information for decision making gained from achievement testing will depend upon the willingness of the educational community to instruct itself in accurately interpreting test results and using the results to the

Table 3

Item Analysis for Identified Priority 1
Objective "Figurative Language"

	Percent Correct Difference Between Local and National	Priority Rating
Objective 40 Figurative Language	-12	
Item 38 Simile	-19	3
Item 53 Simile	-23	1
Item 54 Simile	-20	2
Item 53 Metaphor	-1	4
Item 40 Onomatopoeia	0	5
Item 39 Personification	+6	7
Item 41 Personification	+5	6

benefit of the profession and the ultimate benefit of the individual student.

Talk to the Professionals for School Financing

WILLIAM MARCH
President

FRANK E. WILLIAMS
Executive V.P.

PATRICK H. RENSCH
Vice President and Counsel

HARRY E. COE
Vice President

ROBERT E. ROH
Vice President

WM. (Bill) ABTS
Assistant V.P.

THOMAS B. MEHL
Assistant V.P.

C. W. POORE, JR.
Sec.-Treas.

WAYNE A. RASMUSS
Assistant Sec.

MIKE MULLEN
Representative

MICKEY KRUPINSKY
Representative

DAVE DUNN
Representative

JOHN H. CRAVENS
Representative

Ask us for the list of school districts where we've helped with building expansion. We've been in the business for 40 years and we have a very impressive list of satisfied clients. We understand the election problems and can help with planning and arranging the financing. If you're thinking about building a new school or adding to present facilities, talk to us before you make a move.

SCHWESER SERVICE

. . . The Professional Approach

ROBERT E.
Schweser Company
INCORPORATED

208 So. 19th Street
Omaha, Nebraska 68102
(402) 344-4611
NEBRASKA TOLL FREE 800-642-8438

JAMES V. MITCHELL, JR.

For more than 40 years, the Buros Institute of Mental Measurements has provided a service to educators through candid reviews of commercial tests available for use in assessing human knowledge, attitudes, and behaviors. In this article, the Director of the Buros Institute of Mental Measurements uses a conversational tone to describe to school administrators how the publications and resources of the Institute can be used in review of available tests and selection of a test or testing program.

A HELPING HAND IN TESTING: THE MENTAL MEASUREMENTS YEARBOOK

Perhaps one of the few benefits of getting older is the perspective it gives, or should give, to the practicing administrator. Most of us feel that our experience has provided a rich and varied background for alternative ways of viewing things, defining your choices, and exercising some wisdom in making the right choice. That same experience may also make us especially aware of some themes that seem to be constantly recurring. As an administrator for many years I was particularly impressed with two principal truths about academic leadership: (1) of the tremendous number of choices an administrator makes, each one of those choices involves value decisions and each goes one step further to define what you truly are as a professional and a human being, i.e., as you make your choices so do you also fashion your own personal and professional character and value system; and (2) so many of our concerns as administrators seem to involve questions about quality and effectiveness, either in people or programs. In the face of these persistent, nagging issues about quality, the evidence often seems inadequate to the task. I was always looking for better evidence.

Confronted with these two themes of administrative experience, I found that over time I was making more and more decisions that would place myself squarely in the ranks of those who stood for quality and improvement in academic programs, and also encouraging, persuading, urging, cajoling, wheedling, and doing whatever else was necessary to motivate people to seek better and better evidence of the quality of their programs in terms of student learning and development. I don't think my experience is at all alien to many of you who now read my words.

In the current zeitgeist effective evaluation is mandated not only

James V. Mitchell, Jr. is Director, The Buros Institute of Mental Measurements, The University of Nebraska-Lincoln, Lincoln, Nebraska.

by professional values but also by public demand. Administrators who do not need the public demand and have ready answers to questions about program quality soon develop credibility problems. If there are not equally knowledgeable about the characteristics of their students, they are also suspect. Measurement and evaluation in this context becomes more than a body of esoteric content covered in a university course by that name; it becomes a matter of professional pride and even survival.

In the face of these challenges the typical administrator has had the training needed to work with others to establish a testing program that will provide the required answers. If it is done well, it is not an easy job. In the remainder of this article I shall discuss a particular resource that the capable and well-trained administrator should find invaluable in the development of an effective testing and evaluation program: the **Mental Measurements Yearbooks.**

There are certain principles that are basic to any testing and evaluation program. For effective evaluation it is assumed that an academic program will be structured by a set of objectives at each grade level of the program and that content will be selected in accordance with those objectives and the developmental level of the students in question. It is further assumed that an effective testing and evaluation program will consist of some locally constructed tests and some commercially prepared and published tests, and that the determination of which is used for what is a local issue that can only be resolved by analysis of the objectives and content of the program and the kind of evidence that is wanted. But most teachers and administrators are going to want some kind of outside, external evidence of the status of their students and the quality of their program, and they will look to the commercially published tests for much of that evidence. That is where the **Mental Measurements Yearbooks** can make such an important contribution.

As the new Director of the Buros Institute of Mental Measurements and Editor of the **Mental Measurements Yearbook** I shall not even pretend to present an unbiased view of what the **Yearbooks** can contribute to an understanding of the commercially published test. For over 40 years my predecessor, Oscar K. Buros, undertook a one-man crusade to provide test users with critical evaluations of commercially published tests so that they could choose the best test for the particular purposes they had in mind. The **Yearbooks** have become classic reference works in their field.

How can the busy administrator use the **Yearbooks** most effectively? In some cases he will make use of them himself, and in other cases urge others to make effective use of them. Let us say that as an administrator you are interested in selecting the best achieve-

ment test battery available to assess the academic progress of students in your elementary program. Knowing that involvement of teachers in the selection process can increase their interest in and identification with the testing program, you have appointed a committee to study your needs and make a recommendation. You make available to them all relevant curriculum materials, including statements of objectives, and also copies of **The Eighth Mental Measurements Yearbook** and **Tests in Print II.** If you don't have copies of these you make sure that they know where to get them and how to use them.

The teachers on your committee have some professional savvy, and so they first take a careful look at the objectives and content of your program. With these firmly in mind, they examine the listing of Achievement Batteries in the "Scanning Index" at the back of Volume II of the **8th MMY.** Their appetites thus whetted, they assign different people on the committee the responsibility of reading about and reporting on the different achievement batteries that are reviewed in the **Yearbook.** All of them are made particularly aware of the very great importance of careful matching of the objectives and content of their curriculum with the objectives and content of the achievement battery in question.

As teachers carry out their assignments, they become so familiar with what **The Eighth Mental Measurements Yearbook** has to offer that it soon assumes the character of a long-term and valued friend. They notice that each test entry contains very detailed information about author, publisher, grade levels, forms, content or skill areas covered, manuals, technical reports, costs, teacher's guides and other helps, format, scoring, norms, cost of specimen set, and every other conceivable fact that one would need to know about an achievement battery. They note the listing of references involving studies that have made use of that particular test. But most important of all, teachers take special notice of the critical reviews of the tests or test batteries by highly qualified and well respected people in the field. They expect the reviews to be informative and instructive, and they are. But what they don't quite expect is that they are as forthright as they are; they just don't "pull any punches," as the saying goes. One reviewer of an achievement battery, for example, reports disgustedly that "The multilevel batteries have been diabolically designed to prevent hand scoring by teachers or anyone else" and then later follows up with "The upshot of this Rube Goldberg hand-scoring contraption is that test coordinators will either beg [the publisher] to score the answer sheets or they will opt for another publisher's achievement battery." What a blessed relief from the whitewash and baloney that we so often get!"

As our teachers continue to read, they continue to learn. Most

reviews pay particular attention to the reliability and validity of the test in question, and our teachers relate this to their concern about the battery having curricular validity for the objectives and content of the academic program in their school. They discover through reading of the reviews that each achievement battery seems to have its own particular plusses and minuses, and that the final decision about which battery is best will be a complex one that will have to take all this into account. They take notes on what has been read and prepare to report back to the committee.

The reports back to the committee are an interesting sharing experience. Some of the teachers have been curious about some other tests they have heard about, and they have checked these out as well. In some cases they have not found what they wanted in **The Eighth Mental Measurements Yearbook** and have consulted **Tests in Print II,** which serves as an overall index for all the yearbooks, and have then been referred to earlier yearbooks. Their attitude now is a healthy mixture of determination to find the best test battery for their purposes and a frankly critical posture toward the claims of any test author or publisher. Professionals they are; patsies they're not.

After the reports are made there is a discussion about next steps. Three major decisions are made. First, they will draw up a set of criteria that will help them to determine which achievement battery is best for their purposes. Second, they will narrow down the list of achievement batteries to those three or four batteries that appear to be the most promising. Third, they will order specimen sets of each of these batteries so that each may be considered with the attention, care, and analysis that the decision demands. Since all the information for ordering the specimen sets is contained in the test entries and Publishers Directory and Index of **The Eighth Mental Measurements Yearbook,** the task is easily done.

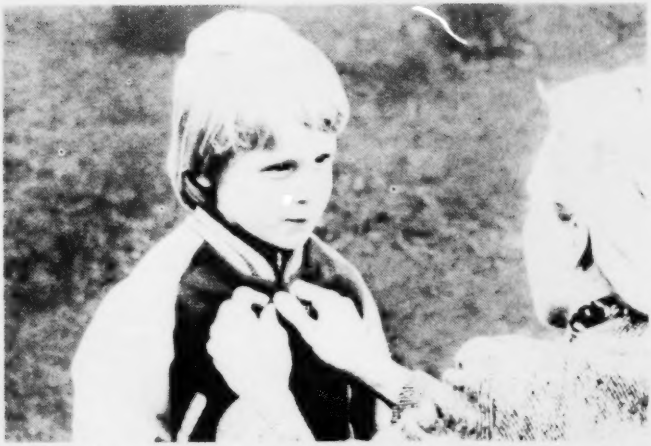
When the specimen sets arrive, they are circulated among members of the committee. Each member of the committee ranks each test battery in terms of the criteria developed. The committee then meets to discuss the batteries and compare the ratings. After some further discussion the committee ranks the three or four test batteries in order of preference and assigns one of their members the task of writing the final recommendation of the committee and summarizing both the process and the rationale that led to the recommendation. After approval by the committee the recommendation is forwarded to the administrator who initiated the task and who is the recipient of the final result. There has been some time and effort invested, but the final product has been a test battery well-selected for a particular set of purposes and a useful byproduct in the form of some on-site inservice training and some teachers

committed to a testing program. It is well worth the effort—particularly when resources like **The Eighth Mental Measurements Yearbook** can be of such help in the process.

For the sake of simplicity we have restricted this discussion to a committee focusing on the selection of an achievement battery. Even more relevant, perhaps, would be a committee recommending an entire testing program, including tests of scholastic aptitude, specific aptitude, vocational aptitude, achievement, interest, personality, and assessment of the learning environment. The **8th MMY** includes all of these and more, encompassing some fifteen different categories of tests and covering virtually every commercially published test in the English language.

It will pay large dividends to any administrator to become thoroughly conversant with what the **Mental Measurements Yearbooks** can offer as a resource to successful testing practice. They are great for browsing. Thumb through and read some of the entries that interest you. Enjoy the no-nonsense approach of most of the reviewers. Become familiar with the six indexes at the back of Volume II. Take a look at **Tests in Print II.** Encourage other people, administrators and teachers alike, to do the same thing. Administrators stand for quality and improvement. They also stand for obtaining the best possible evidence of what students are like, where they are, and how well an instructional program is doing the job. Administrators need a quality product for selecting the best possible tests for obtaining that evidence. **The Eighth Mental Measurements Yearbook,** and its seven predecessors, are that quality product.

We help you take care of what's important.



Taking care of the ones you love is the most important job you have. Protecting them from troubles you can see — preparing them for what you can't foresee.

That's what Blue Cross and Blue Shield of Nebraska is all about. We provide group health insurance for more Nebraskans than any other health insurance company. And we offer the added security of group dental insurance.

To thousands of families like yours, Blue Cross and Blue Shield of Nebraska means they're protected — prepared — secure.

We're kind of proud of that.

Nebraska Dental Service Corporation

Blue Cross Blue Shield of Nebraska

HELPING CONTROL ALL YOUR HEALTH COSTS.

MARDEL R. KOLLS

As public pressure for competency testing programs increased during the past decade, educators became more aware that most approaches to mental measurement were inadequate as a basis for holding learners accountable for classroom learning. Most standardized tests did not provide a match between the tests and curricular or instructional objectives. Another weakness was the tendency of test reporting to focus on normative data without adequate attention to data about individual skills. In this article, the use of the Rasch model, as adapted for use with proficiency tests, is described.

EVALUATING STUDENT ACHIEVEMENT: NEW TECHNOLOGY APPLIED TO OLD MEASUREMENT PROBLEMS

Advances in the use of mathematical models and criterion-referenced testing technology are having profound influences on the ways educational tests are constructed, interpreted, reported and used. Most of the models are based on latent trait or item response theory. One such model, the Rasch model, has been of particular interest to the measurement community as well as educational practitioners. Considered a recent technological advance within the field of measurement, the Rasch model has permitted test developers to overcome many of the difficulties associated with score interpretation and test construction under traditional measurement methodology.

Developed in the early 1960's by George Rasch, a Danish mathematician, the model was developed for use with testing programs for the military. The model has more recently been adapted for use with proficiency testing by Benjamin Wright (1967), University of Chicago. In the fall of 1978, the Los Angeles County Test Development Center (TDC) adopted the Rasch model for competency-based test construction and analyses and established item banks in the basic skills according to Rasch methodology. This article includes a discussion of Rasch methodology as it applies to current district testing needs, and how the Test Development Center has advanced the technology of measurement to meet the increasing demand for alternatives to traditional measurement systems. (Because space will not permit more than a brief description of the Rasch model and how it works, the reader is encouraged to seek additional information on the technical aspects of the

Mardel R. Kolls is Consultant, Test Development Center, Office of the Los Angeles County Superintendent of Schools, Los Angeles, California.

model from other sources. One good source of information can be found in the Summer, 1977 special issue of the *Journal of Educational Measurement.* This issue is devoted entirely to latent trait theory. Additional source materials are included in the list of references at the end of this article.)

New Concerns and Old Measurement Problems

Local districts attempting to implement a state mandated competency testing program became increasingly aware that present approaches to mental measurement fell short in a number of ways. First, if children are going to be held accountable for classroom learning, then a test must be designed to match those learning experiences. Competency tests must be designed to match those learning experiences. Competency tests must be curriculum-based with measurable objectives directly related to existing programs. Standardized tests did not always provide a match between the test and program objectives. These tests also yielded normative data that was not appropriate for assessment of individual skills.

Second, if the intent of a competency-based testing program is to provide a diagnostic assessment system for planned intervention, then districts need test scores that reflect an *objective* measure of student performance. Objective in the sense that performance is not defined by others who have taken the test, nor by the specific set of items that comprise the test. Further, the scale used to measure student progress should remain independent of who it measures and with what items. Under traditional measurement methodology, the scale varies. Score interpretation is dependent upon a "norm" established by a sample group, and the particular set of items administered. Change the sample groups or the set of items and the individual score interpretation changes.

Third, teachers and others working closely with children need a system that is fair for all students, and a system that will provide accurate data on individual student achievement in relation to a specific set of skills. They need a system that can position a student in relation to a total curricular sequence. Teachers need diagnostic data that goes beyond a raw score, and district administrators need normative information for board reporting, decision-making, resource allocations and assessment of overall program effectiveness.

Many districts lacked the technical expertise and financial resources necessary for item development and test construction. Concerned over the legal defensibility of a "homegrown" measurement instrument, districts looked to county agencies for assistance. Responding to local district needs, the Office of the Los Angeles County Superintendent of Schools established the Test

Development Center (TDC). Designed to assist districts in the development of curriculum-based criterion-referenced tests, the center adopted the Rasch model as a basis for test construction and analyses of test results. Calibrated item banks were established in reading, mathematics, language and functional transfer skills (basic skills applied to real life situations).

Rasch Methodology

The Rasch approach to mental measurement is growing in popularity because it provides a scaling system for measuring achievement which is comparable to that used to measure physical attributes.[1] The model is concerned with just two factors: (1) item difficulty, and (2) student ability. The Rasch model specifies that the probability of a person correctly answering a test item is a function of these two characteristics (parameters), one for the person and one for the item. The person characteristic is whatever variable is being measured (reading, math, language, etc.). The person parameter is usually called **ability,** although that term may also be used to refer to achievement. Ability then refers to the amount of a particular characteristic the person possesses that enables her or him to answer questions such as those comprising the test.[2]

The item characteristic is generally referred to as *difficulty.* The difficulty is defined as the point on the ability scale at which an examinee has a 50 percent chance of answering the item correctly. Thus, items and persons are measured on the same scale. Estimating two sets of parameters and evaluating fit to the model are referred to as *calibration.* Item calibration consists of evaluating the fit of items to the model and estimating the difficulty parameter value for each item. For the Rasch model, test calibration involves determining the ability scores to be assigned to students taking the test. This is done by estimating the ability parameter corresponding to each test score, rather than for each examinee as is done in other latent trait models.[3]

When the data used for estimating the parameters fit the model, Rasch model parameter estimates can be characterized as "specifically objective" or "invariant."[4] This means that estimates of the difficulty of items are independent of the persons tested and items comprising the calibration situation. Users of the model describe its objective qualities as "person-free or sample-free" and "item-free, test-free" measurement.

The entire approach is based on simple assumptions concerning student ability and item difficulty, i.e., an able person is more likely to answer items correctly than a less able person and easy items are more likely to be answered correctly than more difficult items. Thus the odds for a student's success on a particular item depend on the

interaction between the student's ability and the item's difficulty.[5]

Because of the objective quality of the model, use of Rasch methodology facilitates item-banking. Pools of test items can be collected, classified and then calibrated. The process of assigning a difficulty value to a pool of items is referred to as a calibration process. Over 60,000 students participated in the initial calibration process that established the basic skill item banks at the TDC. The calibration process is ongoing as new items are continually added to the existing banks. Items that fail to fit the model are systematically removed.

The Student Report

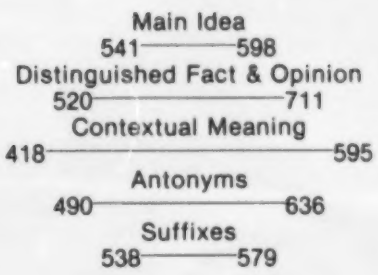
Influenced by the work of Wright (1967; 1977) and his colleagues, the TDC has developed a new approach to reporting individual test results. Because test performance is analyzed in terms of log odds for success, individual scores are reported on a Mastery Level Scale as well as an Achievement Scale.

The Mastery Scale shows the chances out of 100 that a pupil would answer an item correctly, and the Achievement Scale reports the level of student ability in terms of item difficulty. Results are reported as expected or unexpected responses based on the individual's performance on all other test items. In effect, the individual student report is a blueprint of how the student performed on each test item. Based on Rasch methodology, the student's Measure on Scale (score) is always positioned at the 50% mastery level, or that point at which the student would have an even chance of answering an item correctly.

A validity measure referred to as Quality of Measure Index (QMI) is also computed for each student. The **QMI** is based on the student's overall performance and how well he fits the model of getting easy items correct, difficult items wrong. The closer a student's QMI is to 100, the more valid is the measure of his performance.

Skills Continua

The individual reports used by the TDC identify those skills a student can and cannot do, as well as his chances for success or mastery of those skills. To determine where student ability is in relation to total skill development, the TDC developed skills continua for each of the basic skills. Groups of subskills arranged according to calibrated difficulty are displayed graphically to form bands which represent a range of difficulty as defined by the items in the bank. An excerpt from the reading continuum is shown below. Groups of interrelated skills are arranged to form a hierarchy of skills. The unique characteristic of these skills continua is that the information used for development is empirically based on data established by thousands of students. Use of the continua in con-



junction with the individual student report can provide the necessary data for instructional planning.

Reporting Results

Competency-based testing programs in many districts are beginning to bridge the gap between curriculum and evaluation. Rather than allowing the test to dictate the curriculum, instructional programs provide the design for test construction. This very basic change is beginning to have profound effects on utilization of test results.

To assist districts in analysis of program effectiveness based on achievement test results, the TDC has developed a Performance Curriculum Match Report. Based on a discrepancy model, the report form provides information on "expected" and "observed" performance for each objective measured. Deviations between the expected and observed can be both positive and negative, and are reported in terms of percentages. The "expected" percent correct is based on the district's population only (not normative data) and is reported by district, school within district, and grade within a school. The data is specific and facilitates decision-making directed toward program improvement at all levels within an educational system.

The TDC has also developed a Group Analysis Report. This form, designed to facilitate utilization of test results, aggregates the information from the individual Student Reports. By using this form, the classroom teacher can group for instruction by objective enhancing the student's opportunity for measurable gains.

Because this system reports both strengths and weaknesses, teachers are encouraged to work with both "unexpected correct responses," as well as incorrect responses. Hopefully, competency-based testing programs using the Rasch approach will enhance the use of achievement data for planning both challenging learning experiences for the more capable student, and appropriate remedial instruction for the less able student.

Meeting New Challenges

The TDC has continued to respond to the varying assessment needs of educational systems. Most recently, the TDC has developed a Rasch-based handscoring system for use with Special Schools (juvenile court system) and for a CETA project assessment center in Macon, Georgia. Test results can be computed in less than an hour after administration, facilitating efficient placement procedures where group administration and the usual scoring procedures are not possible.

Other projects requiring further development for use of the Rasch model include a pilot testing program of the deaf for Gallaudet College. The center is in the process of designing locater tests to assist teachers of hearing-impaired students in administering the most appropriate test level.

The Test Development Center has also developed a series of cassette filmstrips and pamphlets to assist districts and various educational agencies in articulation of the Rasch approach to mental measurement. (The filmstrips and any additional information may be obtained by writing to: Los Angeles County TEST DEVELOPMENT CENTER, Los Angeles County Education Center, Division of Program Evaluation, Research, and Pupil Services, 9300 East Imperial Highway, Downey, California, 90242.)

Although this system has served to advance criterion-referenced testing technology, it is not the final answer to all measurement problems. However, the Rasch approach has focused attention on development of an objective-based testing system and has provided a means for districts to use test results to benefit the individual as well as the educational system.

NOTES

1. Wright, B.D. "Sample-free Test Calibration and Person Measurement." Proceedings of the 1967 Invitational Conference on Testing Problems. Princeton, N.J., Educational Testing Service, 1968.
2. Rentz, R.R., and Rentz, C.C. "Does the Rasch Model Really Work? A Discussion for Practitioners." **Technical Memorandum No. 67**, Princeton, N.J.: ERIC Clearinghouse on Tests, Measurement and Evaluation, Educational Testing Service, 1978.
3. Ibid.
4. Hambleton, Ronald K. "Putting Latent Trait Models into Perspective: Their Advantages and Disadvantages for District and State Assessment Application." An invited paper presented at the 10th annual Conference on Large-Scale Assessment, Boulder, Colorado. June 9-12, 1980.
5. Rasch, G. **Probabilistic Models for Some Intelligence Attainment Tests.** Copenhagen: Danish Institute for Educational Research, 1960.

ADDITIONAL SOURCES

1. Birnhaum, A. "Some Latent Trait Models and Their Use in Inferring an Examinee's Ability." In F.M. Lord and M.R. Novick, **Statistical Theories of Mental Test Scores.** Reading, Mass.: Addison-Wesley, 1968.
2. Hambleton, R.K. and Cook, L.L. "Latent Trait Models and Their Uses in the Analysis of Educational Test Data." **Journal of Educational Measurement,** 1977, 14, 75-96.
3. Wright, B.D. "Solving Measurement Problems with the Rasch Model." **Journal of Educational Measurement,** 1977, 14, 97-116.

Lincoln Industrial Supply Co.

4600 Cornhusker Highway
Lincoln, Nebraska
402/466-8595
Wats 800-742-7416

Machine Tools & Tooling

Finishing—Honing and Deburring—Turning
Drilling—Grinding—Milling—Cutting
Bending—Forming—Rolling
(Tools for every operation)

Schrader/Bellows Fluid Power

Shop Equipment **Shop Supplies** **Hand Tools**

Which way is your yearbook moving?

Looking for a yearbook printer who cares about *your* book? — one who puts an extra personal touch in getting it together? Walsworth, the world's largest *independent* yearbook publisher, has the modern equipment, plant personnel, kit materials, and sales consultants ready to go to work for you — to make your book the best one ever published.

Bryan J. Smith

Route 9

Lincoln, Nebraska 68506

402-423-2756

A fine way to go.

WALSWORTH PUBLISHING COMPANY

MARCELINE, MISSOURI 64658

JAMES M. BOLAND

Other articles in this issue of **Catalyst** *address concerns which confront school administrators in the design and operation of testing programs. In this article, the author gives school administrators some advice on working with resource personnel who conduct testing of individual students.*

SO YOU WANT TO GET A STUDENT TESTED? TIPS FOR THE BUILDING ADMINISTRATOR

Both state-mandated recommendations for the retesting of individuals already placed in special education programs and a school system's financial situation influence the availibility of psychometrists who can provide individual testing services for the pupils of the school district. Many school psychometrists are responsible for providing service to three or more schools.

What can a building administrator do to accentuate the quality of service available for individual testing of students? The keys to initiating and continuing "preferred service" status for a building revolve around the "four D's." These are Description, Documentation, Detection, and Diversion.

Description. Most psychometricians or central office administrators must prioritize psychological referrals, most of which are marked either RUSH or URGENT. A determination must be made as to which referrals actually are urgent. Psychometrists, in general, do not appreciate labeling and prefer, instead, a description of the child's academic or behavioral performance. The statement that "Johnny must be hyperactive or emotionally disturbed" would not be as helpful as providing selected examples supporting the request for evaluation. An example might be: "Johnny is very restless in class and is rarely in his seat. He has not completed any assignments to date, is extremely disruptive, has very few friends, and requires close supervision in class and on the playground lest he hurt himself or others."

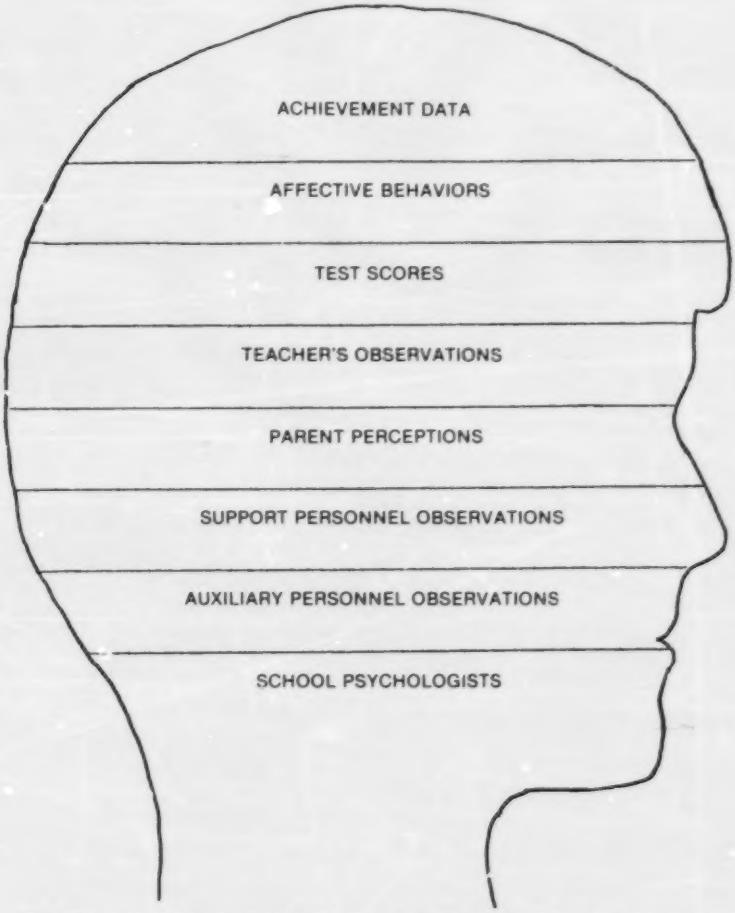
Documentation. A record over a period of time is of primary importance. Documentation of evidence can be either formal or informal and can consist of classwork, report cards, checklists, anecdotal records, rating scales, or participation charts.

Detection. Before a family psychological assessment is undertaken, further detection of the problem should occur. (See Illustration A.) Most students tested by psychometric services are not eligi-

James M. Boland is Assistant Professor of Education, Division of Education, Indiana University Northwest, Gary, Indiana.

ble for any special education program but laws require placement of the student in the "least restrictive environment." Therefore, before placement can become a serious consideration, the school system must have attempted academic or behavioral interventions.

Illustration A



This illustration simply indicates that psychological assessment should be the last resort and should represent only one aspect of evaluation.

Diversion. While diversion of individuals who are, in all possibility, non-placeable lends credibility to referrals, it is perhaps the most difficult practice to "sell" to faculty as the educational burden increases rather than decreases. The fact remains, however, that children in the slow learner category will not be placed in a special education program. Suggestions for teaching these children could be provided by inservice programs which feature specialists in testing, behavior modification, learning disabilities, and related topics. There also are many fine "how to" texts published as an extension of some of the more frequently used standardized instruments. Parent-teacher associations or other community groups could be enlisted to aid in the recruitment of adult helpers whose presence in classrooms can provide teachers with lower adult:child ratios.

This list of the "four D's" is not exhaustive. It is possible to "dazzle" the psychometricians with a knowledge of testing, to "deliberate" with a staff regarding which students should receive top priority in a building, or to "discuss" with parent and child the testing procedure to be used. These are all good ideas. But, the gist of the suggestions offered is to aid the building administrator in creating a favorable working relationship with the individuals who actually prioritize and provide for individual testing of students. The practices of describing, documenting, detecting, and diverting can aid the building administrator in establishing a tone for better communication and for mutual appreciation of professional responsibilities. Such a relationship is beneficial to all concerned.

Bell & Howell APPLE II
Microcomputer System

for education, communication
and problem solving

Bell & Howell Microcomputer System
with Floppy Disk Drive Subsystem

The microcomputer system that opens new dimensions into tomorrow's computer technology is yours today. An advanced system that is easy to use, expandable and surprisingly affordable.
Introducing Bell & Howell Microcomputers. Write or call for full information.

TRIANGLE SCHOOL SERVICE
1409 C Avenue
Sioux Falls, S. Dak. 57104
1-800-843-3784

BELL & HOWELL

JAMES WINDESHAUSEN

A description of the steps taken in one school district to review and revise its testing program and its use of test data is given in this article. Involvement in the review process and the acceptance of a testing program by both faculty and citizens of the community are described as essential to the development and use of meaningful testing programs.

A REVIEW OF A TESTING PROGRAM: THE FREMONT PUBLIC SCHOOLS

In the fall of 1977, the Board of Education of the Fremont Public Schools mandated an indepth study of the practices being used for the assessment of student progress and student instructional needs. There were several desires which led to this decision. The goals of this study of testing programs in the district became: (1) to review information being gathered on students, (2) to review the uses being made of test data collected, (3) to determine how information being gathered was being used to make instructional decisions, and (4) to develop a district response to state and national movements toward competency testing.

The first step taken in organizing for the conduct of the study was the formation of a testing committee. Members of the committee were appointed by the superintendent. Their assigned task was to coordinate the tasks involved in conducting the study and to provide necessary leadership for completion of the study.

While the decision to conduct a thorough review of district testing practices was made during a time of concern about declining test scores and a perceived need for competency testing, the testing committee soon realized that these issues were peripheral to an overriding need for a student-oriented testing program, one that could be used to identify strengths and weaknesses in the performance of individual students so that instruction could be adjusted when there was evidence that skills were not being mastered. This focus on use of test data for improvement of instruction meant that involvement of teachers in the study was critical for any hope of success. More than one hundred faculty members were directly involved in one or more aspects of the study project. Reasons for participation varied but most experienced a sense of involvement and commitment as well as increased understanding of the K-12 instructional programs of the district.

James Windeshausen is Counselor, Fremont Junior High School, and Chairperson, Testing Committee, Fremont Public Schools, Fremont, Nebraska.

While the involvement of faculty from the school district was crucial to the hopes for success of the study, use of expertise external to the district was also a major element in the plans of the Testing Committee. In 1978, a workshop was held to examine competency testing practices used by Westside Community Schools (Omaha, Nebraska); faculty members, members of the board of education, and citizens of the community were in attendance at this workshop. Information gained during the workshop on competency testing served as a springboard for further efforts. Personnel from the University of Nebraska-Lincoln worked with the Testing Committee to help plan subsequent steps in the district's study of testing programs. Much of the work involved in review and development of testing practices was undertaken as part of three field courses offered between 1978 and 1980 by UNL and specifically designed to complete tasks related to the study of testing programs. Summer workshops also were conducted to review information which had been gathered and to set further directions for study and development.

Decisions

Based on input from faculty and citizens, the Testing Committee determined that use of a benchmark approach within a testing program was preferable to the use of test data as a basis for determining whether or not graduation requirements had been met. The concern of faculty and citizens was with ways to help students when test data collected at specified points (benchmarks) in the educational process indicated a lack of mastery of desired skill levels. Content areas which were perceived by faculty and citizens to be most critical were reading, mathematics, writing skills, speaking skills, and consumer education.

To respond to the decision to implement a benchmark testing program in critical content areas, the Testing Committee chose to focus first on reading skills and mathematics skills and to expand at a later point in time to critical examination of skills in writing, speaking, and consumer education.

To establish a benchmark testing program for reading skills and mathematics skills, the first step was to identify the skills being taught and to determine the degree of mastery expected at differing benchmarks. Validation of skills being taught and the expected levels of mastery (skills introduced, partial knowledge of the skills, and mastery of specific skills) was undertaken at each grade level. Once the clusters of skills being taught were identified, it was decided that benchmark testing would be used at grades 3, 5, 8, and 10 to measure mastery by students. At this point in the study of testing practices, several commercial achievement tests were reviewed. From this review, the California Achievement Test (CAT)

was selected as the test which was most compatible with the planned curriculum of the school district. For skills which were not covered by the CAT, teacher-made tests were constructed and, at present, are being validated. The Houghton Mifflin Reading Magazine Tests were used in combination with the CAT to measure skills in reading at benchmarks at grades 3 and 5. Tests for assessment of reading benchmarks at grades 8 and 10 are being constructed by the district.

Concurrent with the development and selection of testing practices for the content areas of reading and mathematics, a recording system was developed to provide students, their parents, and teachers with information about the level of mastery attained. This recording system permits the charting of a student's progress from one benchmark to the next. The level of mastery attained by the student is also recorded. If a student is not measuring up to expected levels of attainment, plans for remediation can be provided. While it is realistic to understand that some students may never attain the level of achievement desired for a specific benchmark, it is equally important that the school should be able to document efforts at instruction aimed at helping the student to attain the desired level of mastery. Information about each student's performance follows the student from one grade level to the next; this makes it possible for teachers, parents and students to see what has been done and to determine what help the student may still require. Other testing information is recorded in the student's cumulative testing folder.

The heart of any testing program is how well it is accepted by both teachers and community. At each step of the review of testing programs conducted by the Fremont Public Schools, inservice programs were developed and implemented so that staff, members of the school board, and citizens of the community were informed and involved. Inservice programs on pilot testing, use of test scores, and how to report to parents were developed and used with staff. Information was disseminated to the community through local news media, service clubs, and parent-teacher associations.

Accomplishments

Although the task begun in 1977 is not completed, the Testing Committee reported a number of accomplishments to the Board of Education of the Fremont Public Schools in September of 1980.

1. A testing program has been developed which is student oriented. The program designed should permit identification of student strengths and weaknesses in skill performance. Recording of student performance at benchmark levels permits maintenance of good documentation of performance and also allows for remediation when necessary.

2. Faculty have been involved in the review of the testing program and in decisions made about new practices. As a result, it is expected that teachers, counselors and administrators will be able to make better use of the test data available when planning for the needs of the individual students.
3. Skills being taught in grades K-10 have been identified for the content areas of mathematics and reading.
4. Testing practices for assessment of mathematics and reading skills at benchmark levels for grades 3 and 5 have been completed. Testing practices for assessment of mathematics skills at benchmark levels for grades 8 and 10 are being piloted. Reading skills for benchmark levels for grades 8 and 10 are being revalidated. It is expected that these testing practices will be in compliance with state guidelines and will replace district use of the Nebraska Assessment Battery of Essential Learning Skills (N-ABELS) tests.
5. A recording system has been developed and implemented which provides teachers, students, and parents with information about student progress toward, or accomplishment of, expected mastery levels.
6. Inservice has been designed and delivered to provide teachers with opportunities to better understand and use data collected through the testing practices of the Fremont Public Schools.
7. Skill books in mathematics for grades K-10 and in reading for grades K-6 have been prepared and provide teachers at all grade levels with information about what is being taught.
8. Materials for use in remedial instruction in mathematics at grades 3, 4, 5, and 6 have been developed.
9. Plans are being made to develop benchmark testing practices for the curricular content areas of writing, speaking, and consumer education.

Conclusion

A testing program which is student-oriented has been developed for the Fremont Public Schools. Benchmark tests to measure student performance in comparison to expected levels of performance permit use of remedial instruction when necessary. The testing program developed is a part of the educational process rather than apart from the educational process—the testing program is one which tests what is being taught. The strengths and weaknesses of skill attainment for each student are identified throughout the student's years in school. The testing program is based on the realization that not all students who graduate from Fremont Senior High School will master all skills at expected levels; however, all students do have a right to know what skills have been attained and what has been provided. The testing program being developed honors these rights.

PAUL J. MALCOM

In the preceding article, a description of the testing program being developed by the Fremont (Nebraska) Public Schools was described. In this article, the comprehensive testing program of the Omaha (Nebraska) Public Schools is described.

THE TESTING PROGRAM OF THE OMAHA PUBLIC SCHOOLS

During the 1976-1977 school year, a testing committee was formed by the Omaha Public Schools to determine how to use tests and test results most effectively. The committee determined that the major goal of testing programs was to provide information that could be used to improve the educational experiences of students enrolled in the school district. The testing program, it was felt, should have benefits for both educators and students.

A good testing program should enable educators to diagnose student strengths and weaknesses, group students for instruction, plan instructional activities, prepare for discussions about educational progress with students and their parents, and provide information to the community and to external agencies. Feedback from a good testing program should enable students to make educational program choices, select educational and vocational goals, and understand their academic strengths and weaknesses.

To achieve the goal and objectives of a good testing program, the design of the testing program was divided into three components: (1) Classroom Planning for Instruction, (2) Audit of Competency (Benchmark), and (3) Aptitude and Achievement Testing. Each is described in the paragraphs which follow.

Classroom Planning for Instruction

The Classroom Planning for Instruction component consists of objective-referenced tests that have been keyed to the curriculum. These tests are usually administered at the beginning of the school year and provide teachers the opportunity to teach to specified critical objectives in the areas of reading, language arts, and mathematics. Students in Grades 4, 5, 6, 7, 8, 9, 10, and 11 are administered objective-referenced tests.

After the tests have been administered, teachers receive a printout that shows whether or not each of the students in their classes has mastered the objectives. Inservice is provided, when requested by the schools, to assist teachers in using the instructional model

Paul M. Malcom, Instructional Research, Omaha Public Schools, Omaha, Nebraska.

that has been designed for this component. The instructional model for the Classroom Planning for Instruction component has six steps: (1) assess student needs, (2) determine critical objectives, (3) organize for instruction, (4) prescribe learning activities, (5) teach to objectives, and (6) monitor student progress.

When the tests are administered in the fall, Step 1 is accomplished. Step 2 is a process where critical objectives are determined if the test contains too many objectives to cover in a single year. For many tests, critical objectives are determined by the district. When this is not done, individual teachers or groups of teachers determine the objectives that should be taught initially. In the Omaha Public Schools, three instructional plans have been devised to accomplish Steps 3, 4, and 5. These are:

1. **Teacher Directed.** When this plan is used, teachers are assigned objectives to teach and students who did not master those objectives are scheduled into the appropriate classrooms. Instructional materials and lesson plans are provided to help teachers plan lessons for the objectives they have been assigned.
2. **Student Directed.** When this plan is used, students study individual cards that deal with objectives they have not mastered. Activities on the cards provide for teacher-pupil interaction. Teachers are encouraged to work with small groups of students whenever possible.
3. **Teacher Directed, Self-Contained Classroom.** When this plan is implemented, teachers use reading or skills groups to facilitate instruction of non-mastered objectives. Instructional materials provided for the two previous methods are also made available for this method.

In each of the instructional plans, student progress is monitored by the administration of followup or interim tests at the conclusion of instruction. After testing, students are informed of their status regarding each of the objectives studied. In most instances, students record their progress in relation to mastery and nonmastery of specified objectives.

Audit of Competency

This component consists of criterion-referenced tests. The objectives upon which the tests are based are basic in nature and narrow in scope. They focus on skills that are considered essential for further learning, especially in Grades one through six. The fourth and fifth benchmarks represent the measurement of the acquisition of basic skills applicable to life situations.

Five benchmark tests are administered beginning at the

readiness level and terminating at the end of high school. Educational measurement focused upon the individual permits early identification of deficiencies so that remediation may take place. Benchmarks 1, 2, and 3 are administered to all students in the elementary grades; Benchmark 4 is administered to designed students in junior high schools; and, Benchmark 5 is administered to designated students in senior high schools. Designated students in junior and senior high schools are those who scored below the 50th percentile on the achievement tests administered in Grades 8 and 10. However, all junior and senior high school students are required to take the social studies subtest.

Benchmark subtests measure skills in reading, language arts, mathematics, reference skills, and social studies. Students who do not meet the established criteria are provided remedial instruction and are retested. Instructional materials have been prepared for use by school personnel in junior and senior high schools. In the secondary schools, students may also enroll in remedial classes that specifically address objectives contained in the benchmark tests. In the elementary schools, classroom teachers provide remediation for their own students. Currently, passing from one grade to another or receiving a high school diploma is not contingent upon successful completion of benchmark tests.

Parents, students, and school personnel are kept informed of the status of individual students concerning each of the benchmark tests. Students are expected to take the tests until successful completion unless they have been officially excused from this requirement.

Aptitude and Achievement Tests

This component includes standardized achievement and aptitude tests and tests that deal with career education. Achievement tests are administered in the spring of the school year to students in Grades 1, 2, 3, 4, 5, 6, 8, and 10. Subtests include reading, language skills, mathematics, and reference skills. Achievement tests contain category objectives that may serve as posttests for some of the Classroom Planning for Instruction tests. Coordinated objective lists have been formulated that show the objectives measured by the tests in each of the three testing components. Scholastic ability tests are administered to students in Grades 3, 6, and 9. A comprehensive career test battery is administered to students in Grade 9.

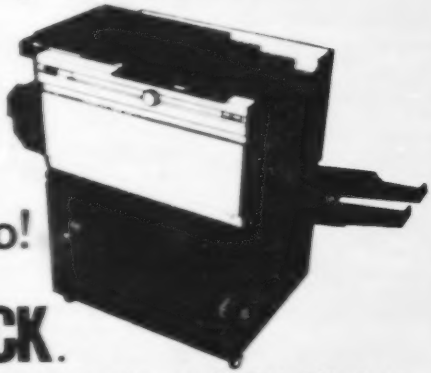
No specific instructional model has been devised for use with this component; however, information about certain schools or the district as a whole is used to indicate changes in curricular emphasis. Each year, a test report is given to the community through

the Board of Education. District percentile ranks are shown for each subtest for each grade level. Also, district scores are shown for the scholastic ability tests.

Summary

The testing program of the Omaha Public Schools consists of a variety of tests divided into three components. Although tests cannot evaluate all of the educational experiences of students, they serve instructional, guidance, and administrative functions. Teachers, parents, and students are provided with information about each pupil's educational progress as measured by the tests.

Why pay 50% more for a copier than you need to!



ABDICK.

The new—and very reliable—A. B. Dick 990 plain paper copier compares to equipment costing 50% more.

The 990 has all of the "most wanted" features of machines costing 50% more—including the versatility to make 11" x 17" copies, copy on both sides of paper, and use most any kind of paper, including overhead transparencies and labels.

Call us today.

NEBRASKA TYPEWRITER CO.

125 NORTH 11TH STREET • P.O. BOX 82081
LINCOLN, NEBRASKA 68501

ROYAL OLYMPIA

TYPEWRITERS

(402) 474-4284

NEBRASKA
ASSOCIATION FOR SUPERVISION
AND CURRICULUM DEVELOPMENT
ANNUAL MEETING

"Evaluation of Curriculum and Instructional Programs"

April 9-10, 1981

Columbus Holiday Inn

Featuring: **Dr. Barbara Day,** President, ASCD
Mrs. Joan Maxwell, University of California Los Angeles

Beginning: April 9th - 6:00 p.m.

Ending: April 10th - 4:00 p.m.

Cost: Members of Nebraska ASCD - $25.00
Non-members of NASCD - $30.00

Reservations: Dr. Jim Walter
Executive Secretary NASCD
104 Henzlik Hall
University of Nebraska-Lincoln
Lincoln, Nebraska 68588
Telephone: 402-472-1988

NCSA-NASCD CALENDAR

DATE	TIME	LOCATION	EVENT
February 11, 1981		Norfolk, Villa Inn	NSASSP Reg III Mtg.
February 13, 1981		Beatrice	NAESP Reg. I Mid-Winter Conference
February 13-16, 1981		ATLANTA, GEORGIA	AASA NAT'L CONV.
February 18, 1981		Omaha, Ross' Steakhouse	NAESP Reg. II Mtg.
February 20-24, 1981		ATLANTA, GEORGIA	NASSP NAT'L CONV.
February 24, 1981	1:00 p.m.	Grand Island, Platt Admin.	NAESP Exec. Bd. Mtg.
March 7-10, 1981		St. Louis, MO	ASCD Annual Conference
March 23-25, 1981		Omaha, Ramada Inn Central	NASBO State Cont.
March 24, 1981	1:00 p.m.	North Platte	NAESP Exec. Bd. Mtg.
March 25, 1981		Crete, Villa Madrid	NASA Reg. I Mtg.
April 4-8, 1981		ANAHEIM, CALIFORNIA	NAESP NAT'L Conv.
April 9-10, 1981		Columbus, Holiday Inn	Nebraska ASCD Conference
April 11-14, 1981		DALLAS, TEXAS	NSBA NATIONAL CONV
April 15, 1981		Norfolk, PSAB	NSASSP Reg. III Mtg.
April 17, 1981		NCSA OFFICE CLOSED	GOOD FRIDAY
April 22, 1981		Bellevue, Bellevue Queen	NAESP Reg. II Mtg.
April 23, 1981		Rushville	NAESP Reg. V Mtg.
April 24, 1981	10:00 a.m.	Lincoln	NAESP Reg. I Exec. Board Mtg.
June 16, 1981	9:30 a.m.	Lincoln	NAESP Exec. Bd. Mtg.
June 24, 1981	9:00 a.m.	Kearney, Holiday Inn	NCSA Exec. Comm. Mtg.
June 24, 1981	1:00 p.m.	Kearney, Holiday Inn	NCSA Retreat
June 25, 1981		Kearney, Holiday Inn	NASA Summer Conf.
June 26, 1981		Kearney	NCSA Golf Tourn.
August 5-7, 1981		Kearney, Holiday Inn	Administrators Days